DOCUMENT RESUME

ED 023 343 HE 000 012

By-Hoyt, Donald P.

The Relationship between College Grades and Adult Achievement. A Review of the Literature.

American College Testing Program, Iowa City, Iowa. Research and Development Div.
Report No-ACT-RR-7
Pub Date Sep 65
Note-61p.
EDRS Price MF-$0.50 HC-$3.15
Descriptors-*Achievement, *Evaluation Techniques, *Higher Education, Performance, *Predictive Validity, Selection

Because grades are the chief means used by college officials and employers to evaluate college performance, their relation to future achievement is significant. Forty-six studies concerned with this relationship were reviewed. The studies were divided into 8 categories--business, teaching, engineering, medicine, scientific research, miscellaneous occupations, studies of successful individuals, and non-vocational accomplishments. Although this area of research is plagued by many theoretical, experimental, measurement, and statistical difficulties, evidence strongly suggests that college grades bear little or no relationship to any measures of adult accomplishment. The findings indicate that 3 major changes in evaluation and selection procedures are urgently needed. First, the meaning of grades should be empirically determined. Second, evaluation procedures in higher education should be drastically altered. Third, these changes should be reflected in policies of selection or acceptance for professional training. (Author/JS)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED023343

ACT

RESEARCH REPORTS

THE RELATIONSHIP BETWEEN COLLEGE GRADES AND ADULT ACHIEVEMENT. A REVIEW OF THE LITERATURE

September, 1965 No. 7

Donald P. Hoyt

HE 000 012

Published by Research and Development Division • American College Testing Program • P. O. Box 168, Iowa City, Iowa 52240

Summary

Research on the relationship between college grades and adult achievement is reviewed. The forty-six studies examined were grouped into one of eight categories--business, teaching, engineering, medicine, scientific research, miscellaneous occupations, studies of eminence, and non-vocational accomplishments.

Although this area of research is plagued by many theoretical, experimental, measurement, and statistical difficulties, present evidence strongly suggests that college grades bear little or no relationship to any measures of adult accomplishment. Consequently, ways to improve the evaluation and selection procedures in higher education are considered.

Table of Contents

Introduction	1
Studies in Business	6
Studies in Teaching	13
Studies in Engineering	21
Studies in Medicine	25
Studies of Scientific Research Contribution	31
Miscellaneous Occupations	37
Adult Accomplishments in Non-Vocational Areas	39
Studies of Eminence	41
Discussion	44
Bibliography	51
Footnotes	57

The Relationship Between College Grades and Adult Achievement.

A Review of the Literature

Donald P. Hoyt[1]

Introduction

What do college grades predict? The question is important because grades are the chief, and often the only, evaluation of the student's college performance. The ultimate consequences of low or high grades are important to the student (who must judge, "Is it worth it?"), to college officials (who must make numerous decisions affecting the student's educational experience), and to employers (who must estimate the professional contribution which the graduate will make). A review of the research on this question raises a number of serious concerns about the relationship between personal characteristics and performance measures and suggests a number of improvements for future research.

Grades are presently important in college because they determine, in large part, the degree and type of educational opportunity which will be available to the student. Nearly all colleges gear their academic probation and dismissal policies to the academic record; students who fail to reach certain standards may be denied the opportunity to continue their studies. In addition, students seeking to transfer to other institutions or to gain acceptance into graduate or professional schools may find their paths blocked by a transcript which contains too many low marks. On the other hand, unusual opportunities are often made available

to students with exceptional grades through honors programs, programs of independent study, or other specially contrived educational experiences. Finally, the omnipresent GPA is commonly used to limit the credit load a student may take, determine his eligibility to participate in extra-curricular activities, certify his qualifications for a loan or scholarship, and recommend him for employment.

Although tremendous effort and expense have gone into the problem of predicting grades,[2] there is a scarcity of studies devoted to the meaning of college grades; a circumstance responsible, in part, for Fishman's recent plea for a moratorium on prediction (Fishman, 1962). While such a moratorium is neither necessary nor practical, Fishman's concern is fitting. We must not be distracted from the basic problems of defining the dimensions of college success and of determining their correlates. Significantly, we must examine, in the light of research evidence now available, whether or not grades can be validly used for their present purposes.

Some Interpretation Problems

In contrast to the literature dealing with the prediction of college GPA, relatively few studies relating college grades to post college criteria have been published, thereby limiting the present review.[3] The complexities inherent in this type of research merit special critical examination.

1. Research in the area has been concentrated on vocational success. Relatively little has been reported in terms of criteria which

might reflect other aspects of academic success (e.g., family life happiness, esthetic appreciation, community leadership, intellectual activities).

2. The range of academic achievement is markedly curtailed, since these studies deal exclusively with college graduates. By definition, subjects in these samples all possessed a degree of academic attainment which their college's faculty judged to be at least minimally acceptable. Many non-graduates achieved below this level. Attenuated correlation coefficients result when a predictor (college GPA) is restricted in range. The amount of restriction varies from study to study, depending on whether employers or professional schools placed a heavy emphasis upon grades in selecting applicants. While this is a source of difficulty in interpreting results, the seriousness of the problem may have been over-stated. Price, Taylor, Richards, and Jacobsen (1963, pp. 105-107q) have provided an extensive technical analysis which suggests that the importance of restricted range is frequently exaggerated.

3. Criterion definition and measurement have constituted a serious problem. For example, salary has been a common criterion. In view of known differences among occupations, companies, and regions, such a measure has obvious limitations. Vocational psychologists (e.g., Super, 1957; Super and Crites, 1962) suggest that work performance should be conceived as a multi-dimensional criterion. An individual may do well on some aspects of his job (e.g., relating to

fellow-employees) but poorly on others (e.g., preparing reports, making decisions). Studies concerned with the relationship of college grades to vocational success are most useful when the complexity of the criterion is recognized and adequate provision has been made for dealing with it.

4. Individual differences among occupational groups, firms within a given occupational group, and colleges produce further complications for the researcher. Common sense suggests that the definition of success in medicine, business, and teaching will require different dimensions. A common criterion, such as salary, neglects differences due to the nation's economic structure and tradition. Similarly differences among firms in their salary and advancement policies produce important but uncontrolled sources of variance. Differing levels of academic ability and grading practices among colleges provide further sources of potential error. This error may be compounded when different departments within a college follow different grading practices or attract students with widely different abilities.

5. Finally, the question of when to assess adult accomplishment is an unsettled and unsettling issue. At one extreme, an immediate follow-up of college graduates might produce negative results because the individual has had insufficient time to establish a reliable record of accomplishment. On the other hand, the greater the time lapse between college graduation and the assessment of adult accomplishment, the more opportunity there is for factors unrelated to the college experi-

ence to affect accomplishment, and the more difficult it becomes to identify relationships between academic and post-college achievements.

Such complexities as these relegate to the future a complete answer to the question, "What do college grades predict?" Nevertheless, educators, employers, and students are forced to interpret the college achievement record as though the answer were already available. These groups might profit from the following survey of studies which have been devoted to this question.

This review is divided into eight sections. Five of these are concerned with specific occupational areas--business, teaching, engineering, medicine, and scientific research. The rest concern a few studies in miscellaneous occupational areas, two studies of success in non-occupational aspects of living, and several studies of eminent men.

Studies in Business

1. Kunkel (1917)

Graduates of Lafayette College from 1876 to 1905 were studied. Ten members of each class were invited to nominate the five most successful members of the class; only 123 of the 300 judges responded. They nominated a total of 301 of the 1593 graduates. Fifty of these were employed in business.

Class rank was determined from college records. For the business sample, 8 were in the upper one-fifth of their class, 9 in the next fifth, and 11 in each of the other three quintiles. There was no relationship between academic standing and success in business.

Comment: The study was done so long ago that it is risky to apply its results to the current scene. No definition of "success" was provided, and the sparse figures on agreement among judges (only 150 of the 301 nominees were named by two or more judges) suggests that idiosyncratic frames of reference were used. It would have been desirable to obtain criterion ratings for every graduate and to correlate these ratings with academic accomplishment; the population to which Kunkel's results apply was of very limited meaning even in 1917.

2. Gambrill (1922)

The 1903 graduates from 11 colleges--Bowdoin, Brown, Dartmouth, Johns Hopkins, Barnard, Goucher, Mt. Holyoke, Smith, Oberlin, the University of Illinois, and the University of Missouri--were surveyed in 1915-16. Just over half of the subjects returned questionnaires

indicating the nature of their present employment and their salaries. Results for men and women were analyzed separately; since most of the employed women were in teaching, only the data for men are reviewed here.

Gambrill computed two correlations between academic record and salary for the 69 business men in her sample. On the assumption that the graduates of the 11 colleges were not different in their achievement, she ranked all 69 subjects on salary and on relative academic achievement. The rank order correlation was .03. In an attempt to control for institutional differences, she computed the correlation between over-all grade average and salary for each college separately and obtained an average correlation (weighted by the number of subjects from each college) of .10. Neither of these correlations was significant, suggesting that, for graduates of these colleges employed in business, there was no relationship between their academic success and their salaries 12 years later.

Comment: Higher education and the business world have changed too much in the past 50-60 years to permit confident generalizations from this study to the present. The sample of colleges was far from random, the return rate poor, and the salary criterion incomplete and potentially misleading since regional and company differences were ignored.

3. Bridgman (1930)

4. Walters and Bray (1963)

These two companion studies were done within a single corporation (American Telephone and Telegraph). Bridgman studied 1310 employees who had graduated at least four years earlier and who had been employed by A. T. & T. for at least half of their professional lives. Walters and Bray studied approximately 10,000 A. T. & T. employees who graduated from college before 1950 and had been employed at A. T. & T. no more than five years after college graduation. The criterion was salary--adjusted for length of service, geographic region, and company department.

In both studies, the statistical analysis consisted of dividing the groups into thirds on the basis of both adjusted salary and rank in class. The results from the two studies were consistent in showing a significant positive relationship between class rank and adjusted salary. For example, in both studies 45 per cent of employees who graduated in the top third of their class earned salaries which were in the top third, while only about 25 per cent of the lowest third academically earned comparable salaries. Correlations were not reported, but it was possible to compute contingency coefficients from the data supplied. These were .37 (Bridgman) and .33 (Walters and Bray), both significantly greater than zero.

Comment: Although salary must be regarded as a limited criterion, the adjustments which the authors were able to make considerably enhance its value. The large samples lend reliability to the findings. The relationships, while not high, are statistically significant and suggest that, at A. T. & T., selection of future employees on the basis of

of their college grades is a useful technique.

The inconsistency of these results with those reported elsewhere in this review raises two questions. First, was it A. T. & T.'s practice to offer higher initial salaries to graduates with impressive transcripts? There is evidence (Brenner & Lockwood, 1965) that initial salary significantly predicts later salary over a long period of time. It would thus be possible to "build in" a correlation between grades and salary. This "self-fulfilling prophesy" may have occurred in a second, but related, manner. It is possible that advancements and accompanying salary increments are based in part on an employee's cumulative record, which includes his college grades. This practice would also produce an artificial correlation between grades and salary. Unfortunately, we could not determine from the reports whether or not these personnel practices existed at A. T. & T.

5. Jepsen (1951)

Male graduates of Fresno State College for the years 1929-1941 were surveyed in 1948. About three-fifths of them responded, including 203 who were employed in business activities. Present (1948) salary was correlated with academic record for these 203 subjects; the resulting r, -.05, was not significantly different from zero.

Comment: Failure to adjust for length of employment may have obscured relationships, particularly since World War II undoubtedly delayed the entry of many late graduates into the labor market. Jepsen implies, however, that analyses not reported in his paper establish

that this was not the case. In addition, limitations of salary as a criterion have already been discussed.

6. Williams (1959)

Alumni of the Stanford Graduate School of Business who had graduated before 1944 and who were located in the San Francisco area were studied in 1958. Salary adjusted for length of time out of college served as the criterion. Among the many predictors were undergraduate grade point average and graduate grade point average. Neither was significantly related to the criterion for this group of 196 men.

Comment: While the criterion was improved by adjusting for length of time out of school and by restricting the study to business men in a single geographic region, the use of alumni from a prestigious graduate school probably produced an unusual restriction in the range of grades and of criterion scores, thus attenuating correlations.

7. Pallett (1965)

This study is the most recent and, in many respects, the most dependable in this section. The sample included 184 graduates of the University of Iowa who had been out of college from five to ten years and who were employed in non-technical jobs in business. As criteria Pallett used ratings of the immediate supervisor. While he obtained an over-all rating (the sum of "Progress" and "Potential" ratings), his major interest was in the specific components of success in this setting. Of the 23 specific characteristics which were rated, 8 made independent contributions to the prediction of the over-all criterion;

these 8 were considered to be elements of "success" in general business. They included Persuasiveness, Drive, Creativity, Leadership, Problem-Solving Ability, Oral Communication, Identification with the Business World, and Identification with the Company. None of the correlations between college grade point average (junior and senior year only) and these elements of success was significant; neither were the correlations between GPA and over-all (Progress and Potential) ratings.[4] The range of these 10 coefficients was from -.06 to +.04.

Comment: While Pallett constructed his rating scales with great care, he was unable to check their reliability. It was necessary for him to assume comparability in the ratings of the various supervisors, a dubious assumption despite his efforts to construct scales with this requirement in mind. By restricting the study to those currently employed in general business, he desirably controlled some variation due to differences among jobs; at the same time, he may have undesirably curtailed criterion ratings since the least successful would probably have terminated their employment before the study was begun. This curtailment would have an attenuating effect upon correlations.

Since six of the ten correlations were negative, the effect of correcting for attenuation would be to make these six more negative. It seems preferable to assume that the criterion restriction was relatively unimportant than that grades were negatively related to effectiveness in business.

Summary of Business Studies

Only the A. T. & T. studies lend any support to the hypothesis that college grades predict future success in business. The weight of the evidence suggests no relationship between the two. Refinements in criterion specification and measurement must occur before conclusive studies can be made. In this connection, the advance by Pallett is noteworthy.

Studies in Teaching

1. Kunkel (1917)

Among the Lafayette graduates studied by Kunkel who were designated successful by their classmates were 55 teachers. Sixty-two per-cent of these were in the upper quintile of their college class; only 5 per cent were in the lowest quintile. Kunkel concluded that there was a direct relationship between academic success and success in teaching.

Comment: While limitations in the criterion and sample have already been noted, there may, in addition, be an important artifact which accounts for the positive finding. Several studies have suggested that students majoring in education are awarded higher grades than those in other academic areas. Kunkel's finding likely reflects this phenomenon. A comparison group of less successful teachers would be necessary to establish a relationship between scholarship and success for Kunkel's sample.

2. Payne (1918)

Graduates of Harris Teachers College (N=144) were rated by their principals after their first year of teaching. Ratings were made on three criteria: management, instruction, and attention to details. Comparisons were made among groups who ranked in the upper, middle, and lower thirds academically. No differences were found on the "management" and "attention to details" criteria. On the "instruction" criterion, 40 percent of the upper third received an "excellent" rating; 27 percent of the middle third and 17 percent of the lower third received

a similar rating. Payne concluded that academic success and success in instruction were positively related.

Comment: Generalization is hazardous, both because of the limited sample (one college) and because of changes which have occurred in education since 1918. This early attempt to deal with the complexities of the criterion problem is laudable even though principles for establishing good rating scales had not yet been established. The positive finding in the area of instruction should be tempered by an apparent non-linearity in the relationship; 10 percent of the upper third received medium or unsatisfactory ratings while only 2 percent of the middle and lower thirds were rated this low.

3. Gambrill (1922)

In her follow-up of the graduates of 11 colleges, Gambrill included 160 teachers--65 men and 95 women. Following the procedures described previously (see "Studies in Business"), she calculated two correlations for each group; one of these was between income and relative class rank, while the other was the average of the correlations between these two variables computed separately for each college. For the men, both correlations were .28 ($P < .01$); for the women, the correlations (.04 and .02) were not significantly different from zero. She concluded that there was, at best, a low relationship between academic success and teaching success.

Comment: The general limitations in Gambrill's study were cited earlier. Possibly there is an artifact in the positive relationship found;

especially if it were the practice of school systems with the highest salary scale to employ graduates with the most impressive transcripts, a not unlikely set of circumstances.

4. Stuit (1937)

School superintendents rated University of Nebraska graduates on seven characteristics believed relevant to effective teaching. On the basis of these ratings, each graduate was assigned to one of four groups--superior, good, average, and poor. A comparison was made between the undergraduate grades of the superior (N=100) and poor (N=46) groups. The former averaged 85.0, the latter 82.4; the difference was statistically significant.

Comment: Omission of the intermediate (good, average) groups dramatizes differences between extreme groups but ignores the majority of teachers. Consequently, the slight difference found would seem to overestimate the relationship between grades and teaching success. Stuit's study, incidentally, confirmed the earlier observation that education majors are awarded unusually high grades; even his "poor" teachers averaged four points higher than the all-university average.

5. Jones (1946)

The sample was composed of 65 Wisconsin graduates of 1941-43 who were teaching in Wisconsin at the time of the study; 57 were women. Two criteria of teaching success were used: supervisory rating (based upon the well-known Wisconsin adaptation of the M-Blank[5]) and pupil gain score (improvement in standardized achievement test scores). Six

academic predictors were examined: freshman-sophomore GPA, junior-senior GPA, four-year GPA, GPA in education courses, grade in the student teaching course, and grade in the educational methods course. Of the 12 correlations computed, only one was significant at the 5 percent level; this was an r of .40 between GPA in education courses and M-Blank ratings.

Comment: The use of more than one criterion is laudable. Unfortunately, pupil gain scores were available for only about half the sample. Interestingly, on that criterion three of the six correlations with grades were negative, though none was significantly different from zero. The one positive finding is suggestive, but it needs to be interpreted in the context of the entire set of studies in this area.

6. Lins (1946)

First year teachers who had graduated in 1943 from Wisconsin were rated by six professional educators, using the Wisconsin adaptation of the M-Blank. Students rated these same teachers, and pupil gain scores on standardized achievement tests were also available for 17 of the 58. These gain scores were adjusted statistically for initial score, intelligence test score, and sensitivity of the instrument to change. Lins used nine measures of academic success which encompassed different types of courses or different periods of college. Each measure was correlated with each of the three criteria.

Eight of the nine GPAs were significantly correlated with the composite M-Blank rating; r's ranged from .28 to .33. No GPA was

significantly related to the evaluations supplied by pupils. Four of the correlations with pupil gain scores, however, were significant, ranging from .52 to .56 for these 17 teachers.

Comment: Lins' study highlights the complexity of the criterion problem in teaching. His three criteria--supervisory rating, pupil rating, and pupil gain score--did not correlate significantly with each other. Though he gave careful attention to the development of the three criterion measures, his use of faculty members as raters suggests that the positive r's with M-Blank ratings may reflect criterion contamination (the faculty raters were likely familiar with the academic records of the teachers). Results on the pupil gain criterion are not subject to this limitation but are based on an extremely small number of cases.

7. Jepsen (1951)

A total of 160 male teachers were included in Jepsen's study of Fresno State graduates. Academic GPA correlated non-significantly (.05) with salary for this group. Although failure to note the number of years of experience seems serious, an index of extracurricular participation correlated .32 with salary for this same group.

8. Erickson (1954)

Nine different criterion measures were obtained on a group of 64 teachers in their second year in Wisconsin high schools. A factor analysis of these measures yielded three factors which were not entirely independent of each other. Erickson labelled these a First Year Rating

Scale Factor, a Second Year Rating Scale Factor, and a Peer-Pupil Response Factor. Ten different GPA's were correlated with each of the three factor scores. None were related to scores on the first factor. The practice teaching grade correlated significantly (.28) with scores on the second and third factor; all other GPA's were independent of these criteria.

Comment: Again criterion complexity is emphasized. Erickson's data indicate that those who have different types of relationships with the teacher disagree in their judgment of his effectiveness; the time at which the judgment is made also appears to be important. The general independence of college grades and teaching success, however defined, was the finding of major interest to us.

9. Jones (1956)

The sample consisted of 46 women who had graduated from Wisconsin in 1951-53 and who were in their second, third,. or fourth year of teaching in Wisconsin high schools. The principal's rating on the M-Blank constituted the chief criterion. Both the professional GPA and the GPA in the major teaching field correlated significantly with these ratings (r's = .29 and .33).

10. Schick (1957)

Like many of the other Wisconsin studies, Schick collected data relevant to this review as an incidental part of his doctoral dissertation. M-Blank ratings were obtained from the supervisors of 72 first year teachers who had graduated from Wisconsin in 1955. The correlation

between the GPA in all professional courses and the M-Blank rating was not significant (r = .05).

11. Massey and Vineyard (1958)

Immediate supervisors, using a five-point scale, gave 62 teachers (who graduated from Panhandle A. & M. College in 1954-56) over-all ratings and ratings on 14 more specific qualities believed indicative of successful teaching. College GPA was correlated with each of the 15 scales. No significant relationships were found between GPA and over-all ratings or ratings on 10 of the 14 specific characteristics. Significant r's, ranging from .28 to .38, were found on "mastery of subject matter," "character, standards, ideals," "competence in English expression," and "general culture."

Comment: The use of a criterion instrument of unknown statistical characteristics weakened a study of much potential value for identifying those elements of teaching success which may be related to academic achievement. This study, like most of the others reported previously, deals with graduates of only one college, thus limiting the generalization which can be made.

12. Cole (1961)

An outside interviewer visited 140 teachers on two occasions, rating each on the Ryans(teacher evaluation) Scale. Subjects were all graduates from an unidentified California college; ratings were adjusted for grade level and experience. An average of the two Ryans Scale ratings correlated .19 with college GPA. This finding was incidental

to Cole's major finding that personality data collected in college correlated .65 with the same criterion.

Comment: The finding that teaching success was much more closely related to personality characteristics than to academic achievement supports a commonly held hypothesis. Frequently, factors such as "personality," "politics," or "luck" are believed to be more important than grades as determinants of success.

Summary of Teaching Studies

Although teaching effectiveness has been studied more frequently than has success in other areas, adequate specification and measurement of criteria remain a central problem. Clearly the solution of this problem will require the collection of many types of evaluative data. Hopefully there will be less future stress on "over-all effectiveness" and more efforts to measure performance in relatively specific terms, as well as maximum use of various sources of judgments--supervisors and peers in addition to pupils.

Only isolated examples from past research indicate a correlation between grades and a measure of teaching success. In those instances where positive results were found, the relationships were generally of a very low magnitude.

Studies in Engineering

1. Rice (1913)

Graduates of Pratt Institute reported their salaries four to six years after gaining their engineering degrees. Correlations between college grade average and salary were computed separately for the mechanical and electrical graduates in each of three classes. The range of correlations was from .16 to .46; two of the six were significantly greater than zero, as was the weighted average of the six (.27).

Comment: Despite the age of the study, the limitations of salary as a criterion, and the relatively small number of graduates from one college, the study seems satisfactory. By computing correlations separately for each class and for both types of majors, Rice instituted some desirable controls which more recent studies frequently overlook.

2. Gambrill (1922)

Only 20 engineers were included in Gambrill's sample of graduates from 11 colleges. As described before, she computed two correlations for each occupational group, one which ignored differences among colleges and one which treated each college separately and obtained an average value. These two methods yielded similar results for the engineering group as for the occupational areas reported earlier. The correlations between rank in class and salary were -.22 and -.23, neither of which is significantly different from zero.

Comment: General features of this study were discussed previously. The number of cases in her engineering sample was extremely small.

Other than the fact that she used different colleges and a longer follow-up than did Rice, the studies were of similar design but produced dissimilar results.

3. Beatty and Cleeton (1928)

Ninety engineering graduates from the 1923 and 1924 classes at the Carnegie Institute of Technology were followed up in 1927. Two criteria of occupational success were used; salary and a rating on the importance of present position. Scholastic standing correlated .03 and .08 with these criteria; neither correlation was significant.

Comment: No information was supplied to permit an evaluation of how adequately the "importance of present position" was measured.

4. Pierson (1947)

Graduates of the School of Engineering at the University of Utah from 1932 to 1941 were studied. The faculty member "best qualified to evaluate his particular accomplishments" rated occupational success on a five point scale. Ratings were obtained for 320 of the 463 graduates. Engineering GPA correlated .43 with these ratings, leading the author to conclude that scholastic achievement was a valid predictor of success in the practice of engineering.

Comment: The criterion ratings were probably made by the same individuals who had earlier judged the academic success of the students. Thus predictor and criterion measures would be contaminated, making tenuous any conclusions about their relationship. The relatively high *r* (.43) is of special interest, however, since it suggests that the attenua-

tion due to the restricted range of academic achievement is not so great that correlations of "respectable" magnitude are unattainable.

5. Martin and Pacheres (1962)

The salaries of 99 engineers employed in a Hughes Aircraft Company research laboratory were compared with their college grades. A barely significant r was obtained for those with four years of experience; no correlation was found for those with six or eight years of experience or for the total group.

Believing that differences among colleges may have confounded the relationship, the authors grouped colleges into "superior", "average", and "inferior" categories. A weighted score was computed for each individual which took into account the reputation of his college and his scholastic record. These weighted scores did not correlate significantly with salary.

Comment: This study is probably the most dependable one in this group. Differences in occupational duties and in companies were controlled. While salary is a more meaningful criterion when these differences are controlled, no single measure is likely to reflect all performance differences. The control for differences among college reputations is worthy of note; however, it constituted a source of error to the degree that reputations were undeserved.

Summary of Engineering Studies

Four of the five studies used salary as a criterion; the weight of the data suggests that it is unrelated to college grades. The other study used

a criterion which appeared to be seriously contaminated. Until more intensive work is done to devise suitable criteria of engineering success, the relationship of college grades to engineering performance cannot be established definitively.

Studies in Medicine

1. Kunkel (1917)

Included in Kunkel's sample of "most successful" Lafayette graduates were 29 physicians. About one-fourth of these finished in each of the first three quintiles of their class; 14 percent were in the lowest quintile. The study is of value primarily for its historical interest.

2. Gambrill (1922)

A total of 30 physicians were included in Gambrill's follow-up of the graduates of 11 colleges. Correlations of class rank with salary were computed by the two methods described earlier. The obtained r's, -.30 and -.20, were not significantly different from zero for this small sample.

3. Peterson, Andrews, Spain, & Greenberg (1956)

A carefully chosen sample of 88 North Carolina general practitioners were intensively observed in practice by a qualified internist. An elaborate record was made of performance during a three-four day period, with separate ratings on six elements of general practice (clinical history, physical examination, use of laboratory aids, use of therapeutic measures, preventive medicine, and clinical records). Combining these judgments constituted the over-all effectiveness rating.

Ratings on this criterion were compared with academic rank from medical school. (Thirty-two medical schools were represented, though most physicians had graduated from an eastern seaboard school). Physicians who graduated in the upper 30 percent obtained significantly higher ratings than did those in the lower 30 percent or middle 40 percent;

the latter groups obtained identical means. The coefficient of contingency between over-all rating and rank in medical school was .36. Further analyses showed this relationship to exist only for the youngest group (age 28-35); for older physicians, there was no relationship between success ratings and medical school standing.

Comment: The study is noteworthy for its careful development of a criterion measure and especially for its thorough assessment of the criterion. A good deal of credence must be given to the ratings of skilled judges who made lengthy observation of the physician in practice. It is unfortunate that undergraduate grades could not be studied. The findings suggest that the quality of medical school performance is significantly related to early professional performance; they also suggest that, as time goes by, medical school rank fails to distinguish among effective and less effective physicians.

4. Richards, Taylor, & Price (1962)

A total of 139 members of the Univeristy of Utah's medical school graduating classes of 1955-1958 were included in the sample. Hospital officials had routinely written letters evaluating the performance of these interns. The chief criterion was the combined rating of two judges who independently quantified the hospital evaluations on a five-point scale; the Spearman-Brown reliability of the combined rating was unusually high (.89). An objective measure of "quality of hospital" was combined with this rating to form a second criterion; this measure presumably took into account differences among hospitals.

Four academic measures were correlated with each criterion; these included undergraduate GPA and GPA for each of the first three years in medical school. Undergraduate GPA was not significantly related to either criterion (r's = .06 and .03). Third year medical school GPA was significantly related to both criteria (r's = .33 and .45), while GPAs in the first two years predicted the combined criterion significantly (r's = .21 and .24 for first and second year respectively). Since the best predictions were made from third year grades, and since the third year focuses on clinical rather than academic work, the authors concluded that academic performance and performance as a medical intern are either unrelated or related only slightly to each other.

Comment: Despite the fact that the criteria which Richards and his colleagues employed were less carefully defined and measured than was true in the Peterson (et al., 1956) investigation, the results of the two were consistent. Both found that medical school performance was related to the effectiveness of the early career performance of physicians. Richards provided further empirical evidence that the restricted range of GPAs is not necessarily a major consideration; third year medical school grades were no more variable than were undergraduate grades, yet the two correlated very differently (.03 and .45) with the combined criterion.

5, 6, 7, and 8. The Utah studies (Price, Taylor, Richards, & Jacobsen, 1964; Taylor, Price, Richards, & Jacobsen, 1965; Richards, Taylor, Price, & Jacobsen, 1965; Taylor, Price, Richards, & Jacobsen, in press). This series of studies represents an unusually thorough

examination of the criterion problem in medicine. A sample of about 500 Utah physicians was selected to represent the diversity of medical practice. Four subsamples were developed: full time medical faculty members of the University of Utah (N=102), board-qualified specialists (N=190); urban general practitioners (N=110); and rural-small town general practitioners (N=105). Through structured interviews, directories and compendiums, faculty and alumni records, curriculum vita and bibliography, polled opinions of medical students, medical school departmental chairmen and peers, questionnaires, and official college transcripts, over 200 different measures of performance were collected for each physician. The 80 measures judged to be most relevant for each of the four subsamples were subjected to factor analysis. These measures included undergraduate GPA, GPA in the first two years of medical school, and GPA during the last two years of medical school for all four groups.

Perhaps the most prominent finding was the complexity of physician performance. From 25 to 29 independent factors were extracted in each of the four samples. While some of the same factors were identified in all samples, a number of factors were found which were unique to a given type of medical practice.

Of most importance to the present review was the emergence of academic achievement as a unique factor in each group; that is, academic performance was unrelated to any other dimension of physician performance. Perhaps the most impressive demonstration of this finding came from correlating each of the three measures of academic performance with the other

performance measures obtained in each of the four samples. Only 3 percent of the 849 correlations were significant; 5 percent would be expected by chance. Of those that were significant, there were more negative than positive coefficients. In the technical report of these studies (Price, Taylor, Richards, & Jacobsen, 1963), the authors provide an extensive analysis of the argument that restricted ranges account for their results, concluding that this factor was unlikely to be of much consequence.

Comment: These studies stand out because of their exhaustive inquiry into criterion assessment. However, the criterion measures lacked the credibility of the Peterson (et. al., 1956) ratings since no systematic observation of clinical practice was included. Statisticians may argue with the factor analytic methods employed, and particularly with the treatment of missing data; there may be some concern with the representativeness of some of the samples since the physicians were all from Utah. Such criticisms seem minor in view of the overwhelming consistency of negative results.

Summary of Medical Studies

Recent investigations in North Carolina and Utah have made substantial contributions both to the problem of criterion measurement and to the meaning of college grades. Further research with more representative samples should be done; hopefully this work can combine the elegance of the Utah statistical approach with the credability of the North Carolina assessment procedures.

At this time, medical school grades seem to bear a positive relationship to the early success of physicians. These grades are apparently not predictive of physician performance after the first few years of practice. The evidence suggests that undergraduate grades are unrelated to success in medical practice.

Studies of Scientific Research Contribution

Studies in this area are characterized by their recency and by their relative sophistication in treating the criterion problem.

1. Taylor, Smith, Ghiselin, & Ellison (1961)

The investigators concentrate on determining the dimensions of the concept, "scientific contribution". They collected about 150 preliminary measures on 107 physical scientists at two air force research centers; the sources for these data included ratings from supervisors, laboratory chiefs, and peers, as well as official records, reports, and publications. The list of measures was reduced to 52 on the basis of a study of the intercorrelations. These 52 measures were then factor analyzed, producing a set of 15 factors presumably descriptive of the dimensions of "scientific contribution". It was possible to develop effective measures for 14 of these 15 dimensions.

Correlations were computed between undergraduate GPA and each of the 14 criteria. Only 3 of the 14 correlations were significantly different from zero--productivity in written work ($\underline{r}$ = .27), creativity rating by laboratory chiefs ($\underline{r}$ = .21), and current organizational status[6] ($\underline{r}$ = .19). Among the criteria which were independent of the GPA were quality of research work, originality of research work, scientific reputation, and over-all performance.

Comment: This was an extremely elaborate and sophisticated study. The findings regarding GPA were incidental to the major purpose of the study. Had this been a central question, we could reasonably expect more

information on the possible effects of differences in grading standards and in the intellectual level of graduates at the various colleges. The setting (air force research centers) may be sufficiently atypical of other settings where physical scientists work that generalization is impaired.

2. Taylor, Smith, & Ghiselen (1963)

The authors report a study done shortly after World War II by the National Advisory Committee on Aeronautics (now absorbed by the National Aeronautics and Space Administration). A total of 239 engineers working as research scientists were involved. The group was ideal for testing the hypothesis that academic performance is related to effectiveness of performance in research since the shortage of engineers had forced the agency to employ some graduates with very poor academic records. The range of college GPAs for the entire groups was 1.40 (D+) to 4.00 (A), with a mean of 2.66. The criterion--merit ratings on performance of research duties--was trichotomized; the triserial r with GPA was .06 (non-significant).

Comment: If there is truth in the belief that a "C" at one college is equivalent to an "A" at another, then failure to control for differences among colleges could be an important source of error. It is necessary to assume that all S's were performing the same or comparable research duties. The definition of research duties was somewhat ambiguous; it was described in the report as well above the trained level but below the supervisory level. No report was made on the reliability of the criterion rating. These ambiguities cloud the interpretation of the findings.

3. Harmon (1963)

This study used 347 physical scientists and 157 biological scientists employed in research capacities by the Atomic Energy Commission. All had earned the doctoral degree. S's filled out questionnaires which surveyed their experience, patents, publications, memberships in scientific societies, and self-ratings of their best scientific or technical accomplishment. On the basis of the questionnaire responses three or more members of the National Science Foundation's selection panels made independent judgments of "scientific competence." Ratings were corrected for rater bias and for differences among fields.

S's were grouped by field (physical science, biological science) and by the year in which the Ph.D. was earned (1949-1951, 1952-1954, and 1955-1956). Correlations were computed between the undergraduate GPA in science courses and the composite rating of scientific competence for each of the six groups. These correlations ranged from -.20 to +.14; none were significant.

Comment: By dealing only with Ph.D.'s, the range of undergraduate GPAs was probably drastically curtailed; we can safely assume that the correlations were attenuated by this restriction. However, if a correction for attenuation were applied, it would increase the size of both negative and positive correlations, making the interpretation even more difficult. Harmon, reporting his dissatisfaction with the questionnaire approach to criterion assessment, pleads for more intensive approaches such as that used by Taylor (et. al., 1961).

4. Taylor (1963)

Subjects were engineers and physicists employed in a research capacity at the Navy Electronics Laboratory (N=103) or at the Naval Ordinance Test Station (N=66). The Thurstone equal-appearing interval method was used to construct scales for measuring "research creativity" and "research productivity." Ratings were obtained from both the immediate supervisor and the secondary supervisor; judgments of these two raters intercorrelated .73 and .66 for the two criteria.

Correlations were computed between the ratings and two measures of academic success--four year undergraduate GPA and the GPA for the last two years of college. Unfortunately, college transcripts were available for only 51 S's. Neither GPA was significantly related to productivity ratings, but both correlated with the mean creativity rating (r's = .32 and .35, $P < .05$).

Comment: Possible differences among colleges were, once again, not controlled. Of even greater significance is the possibility that positive correlations with creativity ratings may be spurious. Opportunities to be creative may be assumed to be more available to those with high GS ratings; and GS rating is likely a function of the amount of education. This would mean that the men in Taylor's sample who had graduate training had more opportunity to display creative talent than did those with only a bachelor's degree. The two groups would be expected to be different in undergraduate grades also, since admission to graduate programs usually depends on high grades. We have no way of knowing if this combination

of circumstances did indeed operate to produce an artificially high correlation in Taylor's study; the composition of his sample was such that it could have.

5. Chambers (1965)

By consulting such sources as the roster of the National Academy of Sciences, starred scientists in American Men of Science, and Who's Who, the author developed lists of "creative" psychologists and chemists. Samples of less creative men in these fields were drawn from membership lists to match the creative samples on the basis of age, amount of education, and opportunity to do research.

A total of 213 psychologists and 225 chemists responded to a number of questions, one of which asked for a self-report of undergraduate GPA. Creative scientists in both fields reported higher GPAs than did the matched control groups. The contingency coefficients were .29 and .24 for psychologists and chemists, respectively.

Comment: The control groups differed from the creative groups in terms of their major interests. For example, 50 of the creative psychologists were in the General-Experimental area and 13 were in Clinical or Educational fields; for the control group, these figures were 22 and 49. Thus "interest" may have had a confounding effect. Preferably, official grades should have been used rather than recall, particularly since the median age of the entire sample was 53. One can only speculate on how correction of these difficulties might affect the modest relationships found.

Summary of Scientific Research Studies

In relation to the studies on other areas, this group of five is sophisticated and well performed. A good deal of progress has been made in defining and measuring criteria. While all findings are not perfectly consistent, college grades seem to have no more than very modest relationships to measures of research performance. There is some consistency in the finding that grades and measure of creativity have low positive relationships.

Miscellaneous Occupations

Kunkel (1917) reported statistics for 65 lawyers and 40 ministers in addition to the occupational groups already reviewed. Half of each of these groups of "most successful" men were is the upper two-fifths of their graduating classes, while about one-fourth graduated in the lowest two quintiles, suggesting a modest relationship between academic and occupational success. Gambrill (1922) reported very similar findings for the 51 lawyers in her sample.

Twedt (1948) followed up 350 graduates of Northwestern's Medill School of Journalism; he obtained a rank order correlation of .20 between grades and salary. Though this correlation was significantly different from zero, Twedt concluded that other factors were probably more important in determining job achievement.

A "professions" group (52 doctors, lawyers, engineers) was included in Jepsen's follow-up of Fresno State graduates (Jepsen, 1951). A non-significant negative correlation (-.15) was obtained between college GPA and salary. Jepsen also reported the correlation between college GPA and salary for his combined group of 471 men; this varied among classes from .12 to -.24, with an over-all r of -.01.

Havemann and West (1952) reported relationships between earnings and self-reported college grades for several groups of workers. The sample was chosen to be representative of all living college graduates in 1947. For men, there were slight positive relationships in the business, high professional (doctor, lawyer, dentist, scientist), low professional

(teacher, clergy, artist), and government groups. No relationships were found for women employed in these same categories. The data were presented in percentage form, and it was not possible to compute correlations or related statistics.

Husband (1957) determined the 1956 salaries of 275 Dartmouth graduates of 1926. He computed median income figures for various college GPA categories. Little systematic relationship was found. For example, those who graduatedwith GPAs between 1.70 and 1.89 earned median incomes of $14,250 while those whose GPA was between 2.50 and 2.69 earned $14,375 and those between 2.90 and 3.09 earned $13,125. At the extremes, there did appear to be a relationship between grades and salary; the 14 graduates with GPAs of 3.30 or higher had median incomes of over $20,000 while for the 17 who graduated with GPAs below 1.69 this figure was only $10,625.

Summary of Miscellaneous Occupations

These studies, while less complete and less carefully designed than many of those reviewed earlier, produced findings consistent with the bulk of research in this area. They agree that, if there is any relationship at all between college grades and salary, this relationship is very slight.

Adult Accomplishments in Non-vocational Areas

1. Plasse (1951)

In 1947, Time Magazine collected data on 9046 college graduates; over 1000 colleges cooperated in supplying names and addresses of all living graduates whose last name began with "Fa." Subjects reported their academic achievement in college; they also answered questions about their economic status, their civic participation, their current events information, their social activity (clubs, organizations), and the satisfactoriness of their home life. Correlations of academic achievement with these non-vocational accomplishments ranged from .01 to .07.

Comment: Plasse's study is most notable for its pioneering effort to assess adult accomplishment in areas believed relevant to the purposes of higher education. Lack of evidence regarding the reliability and validity of the criterion assessments weakened the study, as did his reliance on self-reported academic achievement.

2. Mann (1959)

A carefully selected sample of 290 University of Wisconsin graduates of 1949 was followed up 8 years later. Mann's questionnaire yielded criterion measures in four non-vocational areas: social status of the home, citizenship activities, cultural interests, and amount of additional higher education. Total GPA and the discrepancy between senior GPA and freshman GPA were correlated with these four criteria. Only one of the eight correlations was significantly greater than zero; the exception was the correlation of .39 between the total GPA and the amount of additional

higher education.

Comment: The one positive finding can be explained, at least in part, by the fact that admission to post-graduate training usually requires above average undergraduate grades. The failure to find a relationship between college success and the pursuit of citizenship activities or cultural interests seems important since such criteria are frequently cited as goals of higher education. Of course, the measuring devices must be more adequately constructed and a broader sample of college graduates studied before definitive generalizations can be made. The two studies in this difficult area provide little reason to believe that college grades bear an important relationship to adult accomplishments in non-vocational areas.

Studies of Eminence

A series of studies relating the college record to the attainment of eminence were done in the early part of the century. The studies are primarily of historical interest. Generalizations cannot safely be made, both because of the changes in higher education between 1900 and 1965 and because these studies dealt with very small select samples primarily from private men's colleges in the northeastern part of the country.

Dexter (1902) reported a study of living graduates from two New England colleges. Of those who graduated in the top decile, 5.4 percent were listed in Who's Who; only 1.9 percent of those in the bottom half of their classes were so honored.

Several other studies involving a listing in Who's Who have been reported. Nicolson (1915) studied Wesleyan graduates from 1833 to 1899; half of the "honor men" were listed, as were 31 percent of the Phi Beta Kappa's and only 9 percent of the "plain degree" men. Knapp (1966) and Knox (1947) studied Harvard graduates; Knapp used the classes of 1851 to 1900 and Knox used a sample of eight classes graduating between 1880 and 1925. Their results were similar: about 10 percent of plain degree men, 17 percent of the "Cum Laude" men, and half of the "Summa Cum Laude's" were listed in Who's Who.

These studies are not necessarily contradicted by Olson's recent report that the majority of the college graduates listed in Who's Who averaged "C plus" to "B." (Phi Delta Kappan, 1965). The potential pool of "C plus" students is considerably larger than the "Summa Cum Laude"

pool. The Olson study emphasizes an obvious point: high grades are not a prerequisite to eminence.

Following a different approach, Foster (1910) used three judges to select the 23 most successful men from Harvard's class of 1894. Their academic record (average 2.90) was superior to that of a random sample of 23 graduates from the same class (average 2.36). Langlie and Eldridge (1931) selected the top three scholars and the bottom three scholars from the Wesleyan graduates of 1897 to 1916; the class secretaries and a group of five who were familiar with the graduates (and possibly their academic records) judged the success of these graduates. Although the median rating of the bottom three graduates was "average" (2.9 on a 5 point scale), 89 percent of the "top scholar" group received ratings above this level.

In a similar vein, Bevier (1917) asked judges to identify "eminent" and "successful" graduates from Rutgers' classes of 1862 to 1905. About 7 percent of those who graduated in the upper one-sixth of their classes were chosen as "eminent", while 5 percent from the upper one-third were so nominated. Representation in the "successful" group showed the same slight trend: 35 percent of the upper one-sixth and 32 percent of the upper one-third were so designated. At the highest levels of scholarship, the results were more striking; about one-fourth of the "first honor" men were nominated as "eminent" and over half of these scholars were called "successful."

In one of the two remaining studies of eminent men reviewed, Walters (1921) identified a group of 392 eminent engineers on the basis of their

recognition by one of the four founding engineering societies. He found that 46 percent had graduated from the top quintile of their class, 28 percent were in the next quintile, and 4 percent were in the bottom fifth. Finally, Poffenberger (1925) reviewed the academic records of West Point graduates, 1818-1905, who attained the rank of Brigadier General. A total of 32 percent came from the top fourth of their class, 27 percent from the next fourth, 23 percent from the third fourth, and 18 percent from the bottom fourth.

The studies of eminent men in general suggest that there is a relationship between eminent scholarly work and eminence in adult affairs. Those studies which were expanded to include more representative samples of college graduates suggest that the relationship between academic and adult accomplishments is a modest one at best.

Discussion

While the complexity of the research problem and the diversity of the studies render a meaningful synthesis difficult, a summary of the more dependable studies aids to clarify the relationship between college grades and adult achievement. Pallett (1965), for example, found no relationship between college grades and ratings on any of the eight dimensions he found to characterize success in business. The Utah group (Price, Taylor, Richards, & Jacobsen, 1963) found academic success was independent of the other 24-28 performance characteristics of physicians, though grades in medical school appear to bear low positive relationships to their early career success (Peterson et al., 1956, Richards et al., 1962). In the field of scientific research, college grades have generally been unrelated to performance; occasionally low positive relationships have been reported (Taylor, Smith, Ghiselin, & Ellison, 1961; Chambers, 1964). While the studies of engineers have paid little attention to the criterion problem, in the best designed study, Martin and Pacheres (1962) found no relationship between salary and grades even after adjusting for the differences in reputation among colleges. No one study of teaching success merits special recognition; the review of Barr et al. (1961) showed that, using various GPA's as predictors, the median $\underline{r}$ with supervisory ratings was .09 (33 correlations), the median $\underline{r}$ with pupil gain scores was .00 (10 correlations), and the 4 correlations with pupil or peer ratings ranged from .10 to .28. Studies in miscellaneous occupations and in non-occupational areas are also

consistent in showing little or no relationship between academic success and various criteria of adult performance. Studies of eminent men, however, while out of date and frequently poorly done, suggest that the college student at the top of his class is more likely to attain eminence than his less successful comrades, although the relationship, at best, is a modest one.

Obviously, studies relating college success to post-college accomplishment need to be strengthened and expanded. For example, differences among colleges and among work settings must be more effectively controlled. Both criteria and measuring devices for assessing adult achievements must be more adequately defined. Despite these limitations, however, we can safely conclude that college grades have no more than a very modest correlation with adult success no matter how defined. Refinements in experimental methodology are extremely unlikely to alter that generalization; at best they may determine some of the conditions under which a low positive, rather than a zero, correlation is obtained.

This review therefore confronts us with three major implications. First, the meaning of grades needs to be empirically determined. Second, evaluation procedures in higher education need to be drastically altered. Third, these changes need to be reflected in policies of selection or acceptance for professional training.

1. The Meaning of College Grades

Can we conclude from this review that college grades are actually

or nearly worthless? No. To do so would necessitate showing that grades are invalid representations of the type of student development which they are designed to reflect.

Traditionally, higher education is said to have three major purposes: to preserve, pass on, and enrich the cultural heritage. For the undergraduate student, education focuses almost exclusively on transmitting the cultural heritage. The preservation and enrichment of this heritage is left primarily to scholars and scientists, and to formal preservation devices (such as libraries, museums, galleries, and the professionals who manage them). Undergraduate grades are frequently taken, then, as a relative measure of the degree to which the cultural heritage has been successfully transmitted. In layman's terminology, they presumably tell how much the student knows.

Since there is no necessary relationship between what a person knows and what he does with his knowledge, the validity of grades should be established by determining how well they measure the amount of knowledge the student possesses, not by how "successful" the student is in his subsequent enterprises. Used for such measurement, grades may be valid indices of a student's knowledge. Their failure to predict criteria like those reviewed in this paper hardly constitutes a decisive indictment.

In addition, it is commonly asserted that the measures of adult accomplishment or "success" are highly suspect criteria. Such measures often represent direct or indirect endorsements of a materialistic

philosophy which bears little resemblance to higher education's devotion to truth and wisdom. Results reviewed in this paper may even have been "expected," since "success" in today's world is popularly believed to be a more frequent result of the "glad hand" and the "fast shuffle" than the "reasoned plan" and the "informed viewpoint."

Such logic is sufficiently compelling to warn us against the conclusion that grades are worthless. On the other hand, we need not infer that present methods of assigning grades are inherently valid. In view of the widespread criticism that grades are simply measures of general intelligence, that they refelct only superficial knowledge, that "test-wiseness" and sensitivity to instructor biases are significant sources of error, and that the "knowledge" measured is largely transient, we recommend that intensive studies be made to validate how effectively grades measure the transmission of the cultural heritage. And, finally, in the design of such studies, criterion measures should reflect knowledge of a relatively permanent nature and extraneous variables should be carefully controlled.

2. Evaluation in Higher Education

Educational philosophy differs from institution to institution in accordance with differences in charters, facilities, boards, students, and staffs. While most colleges would probably endorse the general purposes reviewed above, many would add other purposes. College catalogs frequently contain statements which imply additional objectives. For example, most colleges profess to perform a "guidance" function,

helping the individual identify his strengths and weaknesses and plan his future accordingly. The development of vocational competencies and of general skills (e.g., interpersonal competency, communication skill) are at least implied purposes at most colleges. Attitudinal and value development are likewise common goals (e.g., to increase "tolerance," "objectivity," "esthetic appreciation," etc.). Yet the GPA is the only assessment which is typically made of educational progress, with the exception of the negative assessment assigned the student who violates moral, ethical, or legal standards.

There is good reason for believing that academic achievement (knowledge) and other types of student growth and development are relatively independent of each other (e.g., Holland & Richards, 1965). In view of this and the multiple purposes which characterize goals of higher education, how can educational progress best be assessed? We suggest these alternatives: (1) encourage instructors to grade on the basis of multiple considerations, not knowledge alone; (2) encourage the assessment of various characteristics and the subsequent substitution of a "profile of student growth and development" for the present transcript of grades. The second is more appealing than the first. If knowledge is relatively independent of other types of educational growth, a measure which combined multiple indices would be undesirably ambiguous.

On the other hand, the development of a profile would, hopefully, result in broader conceptions of "standards." It should help educators

recognize the individual differences which characterize college students and make explicit some of the drawbacks to the "Procrustean bed" approach to education. College students have different potentials and different temperaments; "development" can most meaningfully be conceptualized, then, from the individual's frame of reference. The plea is not to lower standards but to individualize them more; to encourage and stimulate personal development in whatever dimensions it is best expressed. To be concrete, it means we would be willing to "forgive" a student his inability (or unwillingness) to master a foreign language if he manifested alternative signs of personal development (e.g., composed publishable music, developed his potential for leadership). Dramatic changes in both evaluation and programming in higher education would be the inevitable result of broadening our conception of educational development.

The preceding discussion admittedly goes beyond the data now at hand. Its key assumption, that college grades measure only one relatively independent aspect of educational development, has not been thoroughly established. But is seems demonstrably more consistent with reason and research than the alternative supposition that grades are valid measures of "general worth."

3. Selection of Students for Professional Training

There is another, perhaps less controversial, implication for higher education which the present review suggests; namely, the admission of students to upper division or professional departments. The

practice of basing admission to schools of education, business, engineering, or medicine largely or exclusively on undergraduate grades seems indefensible. It is certain that many potential contributors in these fields are denied the opportunity for professional training. These personal tragedies must represent a sizeable loss to society as well.

Curricula for which professional preparation is a primary goal should accept those students whose potential is greatest for making a professional contribution. This will clearly involve a more comprehensive assessment of student characteristics than the transcript of grades can provide. The present review gives little support to the practice of establishing a relatively high "cut-off" in terms of GPA and then considering "other characteristics" in selecting a professional class.

There is an inescapable obligation on the part of the professional department to evaluate the professional promise and preparation of the student. Society must be protected from the incompetent, and the employers of college graduates have a right to know their strengths and weaknesses. College grades fall far short as comprehensive measures of professional promise or competency.

It is hard to be optimistic that selection and evaluation procedures can be effectively changed immediately. The same complexities which plagued the research reviewed in this paper guarantee no easy solutions. Improved procedures are dependent upon research which relates personal characteristics to performance measures. If we hope to advance tomorrow, we must begin this frustrating and exciting work today.

Bibliography

American College Testing Program. ACT technical report. Iowa City: Author, 1965.

Barr, A. S., Worcester, D. A., Abell, A., Beecher, C., Jensen, L. E., Peronto, A. L., Ringness, T. A., & Schmid, J., Jr. Wisconsin studies of the measurement and prediction of teacher effectiveness. Madison: Dembar Publications, 1961.

Beatty, J. D., & Cleeton, G. U. Predicting achievement in college and after graduation. Personnel J., 1928, 6, 344-351.

Bevier, L. College grades and success in life. Educ. Rev., 1917, 54, 325-333.

Brenner, M. H., & Lockwood, H. C. Salary as a predictor of salary: a 20-year study. J. appl. Psychol., 1965, 49, 295-298.

Bridgman, D. S. Success in college and business. Personnel J., 1930, 9, 1-19.

Chambers, J. A. Relating personality and biographical factors to scientific creativity. Psychol. Monogr., 1964, 78, No. 584.

Cole, D. L. The prediction of teaching performance. J. educ. Res., 1961, 54, 345-348.

Dexter, E. G. High grade men in college and out. Popular Sci. Monthly, 1902, 62, 429-435.

Erickson, H. E. A factorial study of teaching ability. J. exp. Educ., 1954, 23, 1-39.

Fishman, J. A. Some social-psychological theory for selecting and

guiding college students. In N. Sanford (Ed.), The American college. New York: Wiley and Sons, 1962. Pp. 666-689.

Foster. Relation between college studies and success in life. Administration of the college curriculum, 1910. Cited by Bessie L. Gambrill, College achievement and vocational efficiency. New York: Teachers College, Columbia University, 1922. P. 7.

Gambrill, Bessie L. College achievement and vocational efficiency. New York: Teachers College, Columbia University, 1922.

Harmon, L. R. The development of a criterion of scientific competence. In C. W. Taylor, & F. Barron (Eds.), Scientific creativity: its recognition and development. New York: Wiley & Sons, 1963. Pp. 44-52.

Havemann, E., & West, Patricia S. They went to college. New York: Harcourt, Brace & Co., 1952.

Holland, J. L., & Richards, J. M., Jr. Academic and non-academic accomplishment: correlated or uncorrelated? J. educ. Psychol., 1965, 56, 165-174.

Husband, R. W. What do college grades predict? Fortune, 1957, 55, 157-158.

Jepsen, V. L. Scholastic proficiency and vocational success. Educ. psychol. Measmt, 1951, 11, 616-628.

Jones, Margaret L. Analysis of certain aspects of teaching ability. J. exp. Educ., 1956, 25, 153-180.

Jones, R. D. The prediction of teaching efficiency from objective meas-

ures. J. exp. Educ., 1946, 15, 85-99.

Knapp, R. M. The man who led his class in college--and others. Harvard Graduate Magazine, 1916, 24, 597-600.

Knox, J. B. Scholastic standing and prominence. Sch. & Soc., 1947, 65, 194-195.

Kunkel, B. W. Standing of undergraduates and alumni. Sch. & Soc., 1917, 5, 549-555.

Langlie, T. A., & Eldredge, A. Achievement in college and in later life. Personnel J., 1931, 9, 450-454.

Lins, L. J. The prediction of teaching efficiency. J. exp. Educ., 1946, 15, 2-60.

Mann, Sister M. Jacinta. Relationship among certain variables associated with college and post-college success. Unpublished doctoral dissertation, Univer. of Wisconsin, 1958.

Martin, R. A., & Pacheres, J. "Good scholars not always best," cited in research section of Business Week, 1962, Feb. 24, 77-78.

Massey, H. W., & Vineyard, E. E. Relationship between scholarship and first year teaching success. J. teach. Educ., 1958, 9, 297-301.

Nicolson, F. W. Success in college and after. Sch. & Soc., 1915, 2, 229-232.

Pallett, J. E. Definition and prediction of success in the business world. Unpublished doctoral dissertation, Univer. of Iowa, 1965.

Payne, G. E. Scholarship and success in teaching. J. educ. Psychol., 1918, 9, 217-219.

Peterson, O. L., Andrews, L. P., Spain, R. S., & Greenberg, B. G. An analytical study of North Carolina general practice. J. med. Educ., 1956, 31, No. 12, 165 pp.

Phi Delta Kappan. Majority in Who's Who not top students. 1965, 46, 441.

Pierson, G. A. School marks and success in engineering. Educ. psychol. Measmt. 1947, 7, 612-617.

Plasse, W. Comparison of factors in achievement in college and adult life. Unpublished doctoral dissertation, Syracuse Univer., 1951.

Poffenberger, A. T. School achievement and success in life. J. appl. Psychol., 1925, 9, 22-28.

Price, P. B., Taylor, C. W., Richards, J. M., Jr., & Jacobsen, T. L. Performance measures of physicians. Final technical report to the Office of Education on Cooperative Research. Contract No. OE-2-10-093, August, 1963.

Price, P. B., Taylor, C. W., Richards, J. M., Jr., & Jacobsen, T. L. Measurement of physician performance. J. med. Educ., 1964, 39, 203-211.

Rice, D. E. A study of incomes of technically trained men. Sci. Amer., 1913, 109, 116-117.

Richards, J. M., Jr., Taylor, C. W., & Price, P. B. The prediction of medical intern performance. J. appl. Psychol., 1962, 46, 142-146.

Richards, J. M., Jr., Taylor, C. W., Price, P. B., & Jacobsen, T. L. An investigation of the criterion problem for one group of medical specialists. J. appl. Psychol., 1965, 49, 79-90.

Schick, G. J. The predictive value of a Teacher Judgment Test. Unpublished doctoral dissertation, Univer. of Wisconsin, 1957.

Segel, D. E. Prediction of success in college. U. S. Office of Education Bulletin No. 15, Washington, D. C., Government Printing Office, 1934. 98 pp.

Stuit, D. B. Scholarship as a factor in teaching success. Sch. & Soc., 1937, 46, 382-384.

Super, D. E. The psychology of careers. New York: Harper & Bros., 1957.

Super, D. E., & Crites, J. O. Appraising vocational fitness by means of psychological tests. New York: Harper & Row, 1962.

Taylor, C. W., Price, P. B., Richards, J. M., Jr., & Jacobsen, T. L. An investigation of the criterion problem for a medical school faculty. J. appl. Psychol., 1964, 48, 294-301.

Taylor, C. W., Price, P. B., Richards, J. M., Jr., & Jacobsen, T. L. A study of the criterion problem for a group of medical general practitioners. J. appl. Psychol., in press.

Taylor, C. W., Smith, W. R., Ghiselin, B., & Ellison, R. Explorations in the measurement and prediction of contributions of one sample of scientists. Report ASD-TR-61-96, Aeronautical Systems Division, Personnel Laboratory, Lackland Air Force Base, Texas; 1961.

Taylor, C. W., Smith, W. R., & Ghiselin, B. The creative and other contributions of one sample of research scientists. In C. W. Taylor and F. Barron (Eds.), Scientific creativity: its recognition and

2

development. New York: Wiley & Sons, 1963. Pp. 53-76.

Taylor, D. W. Variables related to creativity and productivity among men in two research laboratories. In C. W. Taylor & F. Barron (Eds.), Scientific creativity: its recognition and development. New York: Wiley & Sons, 1963. Pp. 228-250.

Twedt, D. W. Study of 350 Medill School of Journalism graduates. Evanston, Illinois, 1948. (Mimeographed)

Walters, R. W. Scholastic training of eminent American engineers. Sch. & Soc., 1921, 13, 322-329.

Walters, R. W., & Bray, D. W. Today's search for tomorrow's leaders, J. coll. Placement, 1963, 24, 22-23.

Williams, F. J. Predicting success in business. Unpublished doctoral dissertation, Stanford Univer., 1959.

Footnotes

1. The bulk of the library work was done by Larry Braskamp, who also assisted in re-working some of the published data. Without his talented and dedicated effort, this paper could not have been written.

2. Segel's review of the subject in 1934 required almost 100 pages (Segel, 1934). More recently, from 1962 to 1964 a single testing program provided multiple regression equations for predicting grades to nearly 600 colleges (American College Testing Program, 1965).

3. Every pertinent study which we could find is included in this review. No doubt some relevant work was overlooked, and there have probably been many unpublished studies to which we had no access. Additional references which could be supplied by readers will be appreciated.

4. As a matter of incidental interest, college GPA was not significantly related to any of the 25 performance ratings made by supervisors.

5. The M-Blank asks the rater to consider the teacher as (1) a director of learning, (2) a friend and counselor of students, (3) a member of a profession, and (4) a member of the community. Each category includes subquestions to further define the category; ratings are made on a five-point scale within each category. An over-all merit rating is also made and is the criterion used in most of the Wisconsin studies.

6. Includes salary, number of supervisees, level of work.

This report is the seventh in a series published by the Research and Development Division of American College Testing Program. Reports are published monthly and mailed free of charge to educators and other interested persons who have asked to be on the special research report mailing list.

Other reports in the series are:

No. 1 A Description of American College Freshmen, by C. Abe, J. L. Holland, Sandra W. Lutz, & J. M. Richards, Jr.

No. 2 Academic and Non-Academic Accomplishment: Correlated or Uncorrelated? by J. L. Holland, & J. M. Richards, Jr.

No. 3 A Description of College Freshmen: I. Students with Different Choices of Major Field, by C. Abe, & J. L. Holland

No. 4 A Description of College Freshmen: II. Students with Different Vocational Choices, by C. Abe, & J. L. Holland

No. 5 A Description of Junior Colleges, by J. M. Richards, Jr., Lorraine M. Rand, & L. P. Rand

No. 6 Comparative Predictive Validities of the American College Tests and Two Other Scholastic Aptitude Tests, by L. Munday